Doctor-in-the-loop system for feature selection and weighting for AFib classification

Chnashkharhik Movsisyan[1]; Sos Agaian[2]*; Svetlana Grigoryan[3]; Evgeny Karpulevich[4]; Arman Darbinyan[5]; Lusine Hazarapetyan[6]

[1]Department of Informatics and Applied Mathematics, Russia n-Armenian University, Yerevan 0051, Armenia.

[2]Department of Computer Science, College of Staten Island (CSI), CUNY, City University of New York, New York, NY 10314, USA.

[3]Department of Arrhythmia, Yerevan State Medical University, Yerevan 0025, Armenia.

[4]Department of Information Systems, Ivannikov Institute for System Programming of the Russian Academy of Sciences, Moscow 109004, Russia.

[5]Department of Mathematics and Mathematical Modelling, Russian-Armenian University, Yerevan 0051, Armenia.

[6]Department of Cardiology, Yerevan State Medical University, Yerevan 0025, Armenia.

*Corresponding Author: Sos Agaian

Department of Computer Science, College of Staten Island (CSI), CUNY, City University of New York, New York, NY 10314, USA.

Email: sos.agaian@csi.cuny.edu

Abstract

Atrial Fibrillation (AFib) is the most common sustained arrhythmia, which causes significant morbidity and mortality among the population. AFib is strictly associated with other heart problems and can drastically increase the risk of transient ischemic attack, strokes, and heart failure. Various common risk factors, such as diabetes, hypertension, smoking, age, obesity, hyperlipidemia, and others, are common in patients with AFib. If someone has any of these risk factors, they are at risk for AFib heart failure, diabetes, or coronary artery disease. Thus, diagnosing early and treating it is essential. The implications of Ischemic Heart Disease (IHD), arterial hypertension (AH), and AFib have not been satisfactorily studied. This article aims to develop an effective Doctor-in-The-Loop machine learning system to enhance the performance of heart disease diagnosis by (a) extracting the most contributed to AFib risk predictors; (b) weighting the extracted features based on their effect on AFib progression; (c) generating a new dataset containing patients with IHD and AH; and (d) classifying AFib using the newly constructed dataset. Computer simulation results on the RIC AFib Dataset show that (1) the Age, Diastolic Blood Pressure, IL-6, TNF-alpha, and TNF-beta1 features have the highest effect on AFib progression, (2) the weighted risk predictors overperformed the AFib classification results reaching up to 98.36% accuracy. Also, the evaluation of the presented method on the UCI Heart Dataset shows that it achieved 91.67% heart disease classification accuracy, which is 6% higher than the Spencer et al. method.

Introduction

Atrial fibrillation (AFib) is the most frequent cardiac arrhythmia in clinical practice [1-8]. It is a significant risk factor for ischemic stroke and provokes a significant economic burden, substantial morbidity, and mortality. Atrial fibrillation is a heart rhythm disorder caused by degeneration of the electrical impulses in the upper cardiac chambers (atria), resulting in a change from an organized heart rhythm to a rapid, chaotic rhythm. People with AFib are 5 to 7 times more likely to have a stroke than the general population average. The Ischemic Heart Disease (IHD, which means that the heart is not getting enough blood and oxygen) causes more deaths, morbidity,

and financial burden than other diseases in western societies. It has been estimated that 33.5 million people worldwide are diagnosed with AFib [1-11]. AFib is associated with a range of cardiovascular conditions such as Arterial Hypertension (AH), Chronic Heart Failure (CHF), IHD, which are proven risk factors for the development, persistence and progression of AFib. Due to the increase in life expectancy improvement in such patients in survival rates, an increase in the prevalence of AFib is currently observed.

The association of AFib with IHD and AH is frequent. Therefore, many factors are possibly involved in the occurrence and development of AFib and its progression [11-20,22-23]. Practitioners are faced with correlating structural and functional features with symptoms such as inflammation and fibrosis.

In general, traditional patient information, management, and therapy systems suffer from an absence of intelligence. They successfully offer basic patient management capabilities to their end-users, but they do not provide substantial decision support functionalities or automation to lend a helping hand to clinicians. Also, the most commonly used AI systems are "humans-out-of-the-loop" ones that cannot understand the context and hence cannot reason about interventions and retrospections [24].

Additionally, medical machine learning approaches suffer from weakly structured and non-standardized training data, quality of features (clinical characteristics), rare events, or dealing with uncertainty. As mentioned in [19], biomedical data sets are full of uncertainty, incompleteness, etc., which entails severe machine learning performance limitations. For example, the dataset may miss essential data or be incomplete, noisy, dirty, imbalanced, or uninformative. Therefore, fully automated approaches are difficult or even impossible to implement, or at least the quality of results from automated procedures might be questionable [19,20]. To solve these problems, one needs a doctor's guidance to influence with their experience the work of the machine learning system, including extracting features, decision support, forecasting, and finally, to move from a single-domain AFib classification to comprehensive description of AFib patients (see review on 2020 European Society of Cardiology guidelines) [24].

Recently, a new paradigm know as "Doctor-in-The-Loop" (DTL) has gained attention in the data-information-driven medicine community. It aims to use knowledge discovery to improve medical treatments with the "human-in-the-loop" concept and ensure that machine learning systems routinize the correct information [17,18].

The next question we will investigate in this article is which features we should use to create a predictive model. Quality of Features (QF) has been a critical issue in machine learning. Effectively distinguishing important and inadequately relevant features in terms of their imapct on particular heart diseases can improve the predictive models' accuracy, particularly in learning performance. Various approaches have been proposed to address this issue for decades. QF is one of the main pieces of feature engineering and can be roughly divided into feature selection and feature weighting algorithms. Feature selection is the process of obtaining a subset from an original feature set according to a specific feature selection criterion. It isolates the

most consistent, non-redundant, and relevant features in model construction. Also, feature selection is increasingly important as the size and complexity of the average dataset continue to grow exponentially [25]. Spencer et al. proposed three methods to select a set of features to improve the accuracy of heart disease classification on the UCI Heart Disease Dataset [1]. The best performing model was the BayesNet algorithm on Heart ChiSq features, which achieved an accuracy of 85%.

The goal of feature weighting is to estimate the relative importance of each feature and assign it a corresponding weight [26,27]. Alternatively, feature weighting approximates the optimal degree of an individual feature's influence with a training set [28]. This study investigates the weight of feature's impact on AFib progression and generates specialized databases.

The goal of this article is to develop a doctor-in-the-loop machine learning system to identify and weigh the risk predictors that contribute to the onset of AFib associated with IHD and AH, as well as perform AFib classification based on weighted predictors.

The main contributions of the proposed method are:

a) A novel "doctor-in-the-loop" AFib risk predictors identification system;

b) A novel dataset drove feature weights adjustment procedure for AFib progression with IHD and AH patients;

c) Evaluation results using the newly constructed dataset and UCI Heart Dataset.

The remainder of the article is organized as follows. A literature review on interest modeling in multiple and single application environments and weighting methodologies is discussed in Chapter 1. The proposed approach and system design are discussed in Chapter 2. Chapter 3 discusses the results. Chapter 4 presents the evaluation of the proposed method on benchmarking the UCI Heart Dataset. Finally, discussions and conclusions for future work are presented in Chapter 5 and Chapter 6.

Materials and methods

Proposed methods

This section introduces the main workflow of the proposed method. Additionally, it presents the methodologies used in the computer simulation.

In this study, we develop a system to identify AFib in patients with IHD and AH using an ML-based computational approach with four primary techniques. These are predated by data collection, which fits into study analysis. The workflow of the proposed study is shown in Figure 1. It can be listed as follows:

Step 1: Expert-driven features (clinical characteristics) selection based on a statistical analysis of correlation coefficients.

Step 2: Risk predictor identification and weighting based on Generalized Linear Model (GLM) results.

Step 3: A new dataset constructed on weighted values of risk predictors.

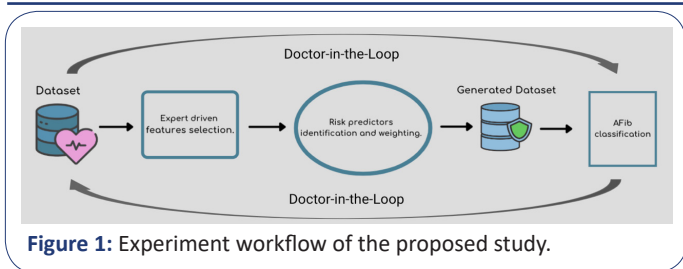**Step 4:** A Fib classification based on a new dataset.

Figure 1: Experiment workflow of the proposed study.

Pearson correlation-based feature selection

Most feature-weighted algorithms do not consider the correlation between features, so the redundancy with interference harms the final classification results. Patients' clinical characteristics usually contain many parameters that may support disease diagnosis [16]. However, sometimes it leads to non-precise results. Hence, we analyzed the factors based on Pearson Correlation Coefficient (PCC; threshold equals 0.4) [29,30]. Sometimes, conclusions drawn from statistical analysis do not include the necessary domain expertise (medical knowledge) to provide the full picture. Thus, the feature selection criterion is based on PCC results and enriched with additional information and doctor know-how. Moreover, the experiments show that the doctor's knowledge on feature selection improves the classification results (Table 3).

Regression for risk predictors identification

The second step of analyzing the clinical characteristics is done using selected features. A logistic regression model was used from the GLM family to identify the features that contribute the most to the disease's progression [31-34]. It models how the "odds" of success for a binary response variable Y depend on a set of predictors. In the case of the logistic regression model, the GLM would have this form:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p, \qquad i = 1, \cdots, n \qquad (1)$$

where $\pi_i = P(y_i = 1), \beta_j$ is a regression coefficient, y_i is an observation, x_i is a set of corresponding predictors, $P(y_i = 1)$ is a probability of ith patients having AFib.

Elastic Net (ENET) that is a combination of l_1 (Lasso) and l_2 (Ridge) penalty functions [46] was used as a regularization method. ENET seeks to find the regression coefficients which can minimize the following representation:

$$\hat{\beta}(ENET) = \underset{\beta}{\arg\min} \sum_{i=1}^{n} \left(y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i) \right) \qquad (2)$$
$$+ \lambda((1 - \alpha)\|\beta\|_{l_2}^2 + \alpha\|\beta\|_{l_1})$$

Where $\lambda \geq 0$ is the tuning (penalty, regularization, or complexity) parameter that regulates the strength of the penalty (linear shrinkage) by determining the relative importance of the data-dependent empirical error and the penalty term. α sets the degree of mixing between Ridge and Lasso. The ENET is equivalent to the l_1 when $\alpha = 1$ and as α decreases towards 0, the ENET approaches the l_2 regression. The automated variable selection is being made by l_1, while l_2 improves the prediction [35,36].

To get the best set of ENET optimal parameters, we plan to use the grid-search with the cross-validation method [36,37]. Grid-search is a widely used method that picks the best model parameters from a list of parameter options for a given optimization problem by automating the 'trial-and-error' method [46]. It is an exhaustive parallel search and will find the best way to tune the hyperparameters based on the training set. Cross-validation is the procedure of training learners using one set of data and testing it using a different group [36]. The inner optimization finds model parameters β, which minimize the training loss L_{Train} given hyperparameters λ and α. The optimization procedure chooses λ and α to reduce the validation loss L_{Val} [38].

$$\underset{\lambda,\alpha}{\arg\min} L_{Val}\left(\underset{\beta}{\arg\min} L_{Train}(\beta,\lambda,\alpha)\right) \qquad (3)$$

Statistical significance was set at p<0.05 to get those risk predictors contributing to the onset of AFib [31,39].

Risk predictors weighting method

As mentioned above, feature weighting aims to estimate the relative importance of each feature and assign it a corresponding weight. Alternatively, feature weighting approximates the optimal degree of individual features' influence with a training set [28,40]. One may find the survey of the feature weighting method's advantages and limitations in [40,41]. The authors mentioned that there are three different ways to use the weight feature: (1) setting weights for all features, (2) setting weight functions for each feature, named as Per-Feature weighting, and (3) setting weight functions for each feature within each class, named as Per-Class-Per-Feature weighting. We will focus on setting weights for the identified risk predictors, which mathematically can be represented as

$$F * W = w_1 f_1, w_2 f_2, \dots, w_n f_n \qquad (4)$$

where risk predictors' weight is $W = \{w_1, w_2, \dots, w_n\}$ and risk predictors are $F = \{f_1, f_2, \dots, f_n\}$. w_i is being produced by a trained regression of GLM according to each feature impact on AFib progression.

The next technique of the study is the construction a new dataset. The dataset contains the weighted risk predictors that impact AFib progression the most. This way, we calculate the weight of each risk predictor based on its effect on disease progression and use the weight of each risk predicor instead of its clinical/laboratory value. This facilitates building a system that will provide AFib classification more accurately. For the classification purposes, five different supervised ML algorithms have been used: Logistic Regression, Gaussian NB, Decision Tree, Random Forest, LDA, and KNN [42,43]. Additionally, the novel system enables precise identification of AFib patients from low to high risks supporting intermediate risk levels. The performance of the constructed dataset is evaluated using some extensions of Receiver Operating Characteristic (ROC) [44-46,47,48].

Novel RIC AFib dataset

We observed 257 patients with IHD and AH hospitalized at the Department of Cardiac Arrhythmias of the Research Institute of Cardiology (RIC) in Armenia and some on an outpatient basis. This implies that RIC Dataset consists of data that is specific to the Armenian local population and its characteristics (atmospheric conditions, geographical location, social lifestyle). Of the total group of patients examined, the study to determine the stratification of risk factors for progression of AFib included 213 patients with paroxysmal, persistent, and permanent forms of AFib (classification ESC 2016, 2020). Forty-four patients with AH and IHD but without AFib similar in gender and age were examined as a control group. The considered classes are described in Table 1:

Table 1: The description of each class.

Class Type	Description
Paroxysmal	It occurs when a rapid, erratic heart rate begins suddenly and then stops on its own within 7 days. It is also known as intermittent AFib, often lasts less than 24 hours, and does not require treatment.
Persistent	It begins spontaneously. It lasts at least 7 days and may or may not end on its way.
Permanent	In people who have had AFib for a long time, the Heart may not be able to return to a normal rhythm.
Control Group	The patient in this group does not have any kind of AFib symptoms.

Patients inclusion criteria were a) systolic blood pressure 140 mmHg. and higher, diastolic - 90 mmHg. and higher (ESC 2018), b) unstable angina (Recommendations for the treatment of stable coronary artery disease ESC 2013), c) the presence of recurrent (paroxysmal, persistent) or chronic (permanent) forms of AFib (ESC 2016, 2020 classification).

The patient examination includes general clinical and additional research methods: hemogram, lipidogram, electrocardiogram, EchoCG, 24-hour Holter ECG monitoring, biochemical blood tests (determination of coagulogram, lipid spectrum, fibrinogen), quantitative determination of hs-CRP, cytokines-IL-6 and TNF levels, as well as a marker of fibrosis - TGF-beta1. Detailed information on the patients with IHD and AH containing the study dataset is described in Table 2.

In this study to identify and weight the rsik predictors, 213 patients with 3 forms of AFib were considered Positive AFib, and the rest (control group) as Negative AFib. After identification and weighting, the classification has been done on described 4 classes.

Results

This section presents the computer simulation parts, introducing characteristic clinical analysis with the doctor's authority, selected features weighting procedure, new dataset construction, and AFib classification.

Expert driven feature (clinical characteristic) selection

Demographic, clinical, and laboratory data of studied ischemic heart disease and arterial fibrillation patients are shown in Table 2. For this study, 257 patients were examined; and the number of clinical characteristics applied to each of them was 27. The suggested methodology for supporting the risk predictors' identification of AFib progression is based on the clinical characteristics collected from patients. Thus, the method of selecting non correlated characteristics as attributes of risk predictors' construction is considered [30,49]. The Pearson Correlation Coefficient (PCC) of all 27 clinical characteristics is shown in Figure 2.

According to the expert analysis and the PCC (a correlation coefficient of over 0.4), the following features have been eliminated: *QRS, HF, LAD, LV EVD, LV ESV, PAP, Gender, FIBR, TIA, Systolic blood pressure, Hypertensive crisis, Pulse, and IVST*. Figure 2 shows that there are clinical characteristics (*LAV, TGF-beta1*)

which have higher correlation than the specified threshold (0.48), but we still keep them for further analysis. This part is confirmed by doctors with medical statements, as *LAV* and *TGF-beta1* characteristics are crucial for AFib progression. The selection performance was evaluated by some classifiers. Table 3 shows the outcome. As we can see in Table 3, the classification results outperform while adding *LAV* and *TGF-beta1 correlated* characteristics.

Table 2: Clinical characteristics of the study patient.

Abbreviations	Clinicopathologic factors	Units	Groups
Gender		%	
Female			121
Male			136
HC	Hypertensive crisis	%	
Yes			127
No			130
TIA	Transient ischemic attacks	%	
Yes			55
No			202
IHD	Ischemic heart disease	%	
Yes			225
No			32
MI	Myocardial infarction	%	
Yes			61
No			196
Age		years	59.09 ± 6.11
SBP	Systolic blood pressure	mm/Hg	161.98 ± 8.05
DBP	Diastolic blood pressure	mm/Hg	96.71 ± 6.46
Pulse	Pulse	bpm	89 ± 11.16
HF	Heart failure	%	0.88 ± 0.81
QRS	QRS complex		102.53 ± 10.58
BMI	Body mass index	kg/m²	30.54 ± 1.96
LAD	Left atrial diameter	mm	40.21 ± 3.13
LAV	Left atrial volume	mL	65.37 ± 10.81
LV EDD	Left ventricular end-diastolic diameter	mm	54.19 ± 2.25
LV EDV	Left ventricular end-diastolic volume	mL	112.76 ± 9.11
LV ESV	Left ventricular end-systolic volume	mL	48.53 ± 9.58
IVST	Interventricular septum thickness	mm	12.7 ± 1.3
LV PVRT	Isovolumetric relaxation time	ms	12.02 ± 0.97
LVWT	Left ventricle posterior wall thickness	mm	0.46 ± 0.02
EF	Ejection fraction	%	48.41 ± 2.8
PAP	Pulmonary arterial pressure	mm/Hg	25.91 ± 7.2
FIBR	Fibrinogen	mcm/l	13.19 ± 2.23
CRP	C-reactive protein	mg/l	4.36 ± 1.85
IL-6	Interleukine-6	pg/ml	26.57 ± 11.58
TNF-alpha	Tumor necrosis factor -alpha	pg/ml	0.43 ± 2.82
TGF-beta1	Transforming growth factor- beta1	pg/ml	708.25 ± 220.0

Classifier Name	Without Expert & Pearson Correlation				With Expert & Pearson Correlation			
	Sensitivity (%)	Precision (%)	F1-score (%)	Accuracy (%)	Sensitivity (%)	Precision (%)	F1-score (%)	Accuracy (%)
Logistic Regression	83.93	82.62	83.17	84.62	86.20	84.25	85.00	86.15
Decision Tree Classifier	82.76	81.00	81.67	83.08	86.25	86.25	86.25	87.69
Random Forest Classifier	81.30	81.00	81.65	83.08	88.37	84.37	84.16	84.62
Linear Discriminant Analysis	86.25	86.25	86.25	87.69	93.07	90.75	91.66	92.31

The performance evaluation criteria are defined in Table 4: where True Positives is the number of correctly classified patients with AFib, True Negatives is the number of correctly classified patients without AFib, False Positives is the number of incorrectly classified patients with AFib, False Negatives is the number of incorrectly classified patients without AFib [42,43].

Table 4: Used metrics.

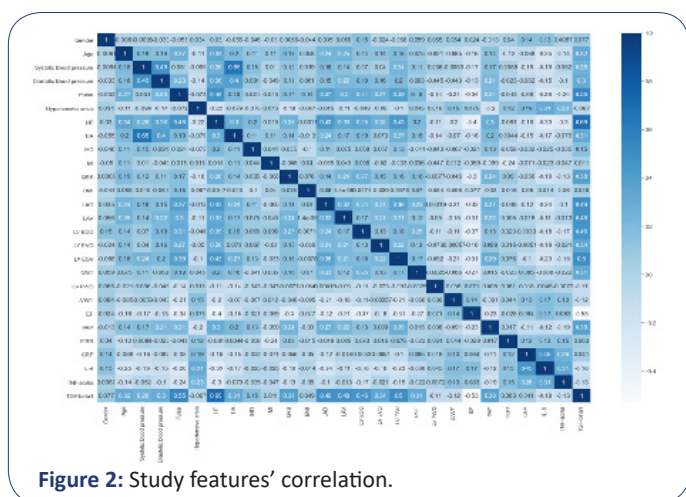| Measure	Formula
Adopted Accuracy	$\frac{True\,Positives + True\,Negatives}{Size\,of\,Dataset}$
Sensitivity	$\frac{True\,Positives}{True\,Positives + False\,Negatives}$
Precision	$\frac{True\,Positives}{True\,Positives + False\,Positives}$
F1 Score	$2 * \frac{Precision * Sensitivity}{Precision + Sensitivity}$

Figure 2: Study features' correlation.

Risk predictors identification and weighting

A Generalized Linear Model was built to find the predictors contributing to the onset of AFib. The selection of tuning parameters λ and α has been done via an optimum value found through the GridSearchCV minimization approach [36]. The ranges for grid search were α values between 0 and 1 with a 0.01 separation and λ values from 1e-5 to 100. The experimental results show that ENET regularization behaves properly when λ and α hyperparameters are set as follows: $\alpha = 0.5$ to provide an equal contribution of each penalty to the loss function, and $\lambda = 0.004$ weight to the regularization. Some results are shown in Table 5.

p<0.05 statistically significant level was considered for risk predictors' selection [31,32,39]. The results according to the best hyperparameters are shown in Table 6. This provides us with five features: *Age, Diastolic blood pressure (DBP), IL-6, TNF-alpha, and TGF-beta1.*
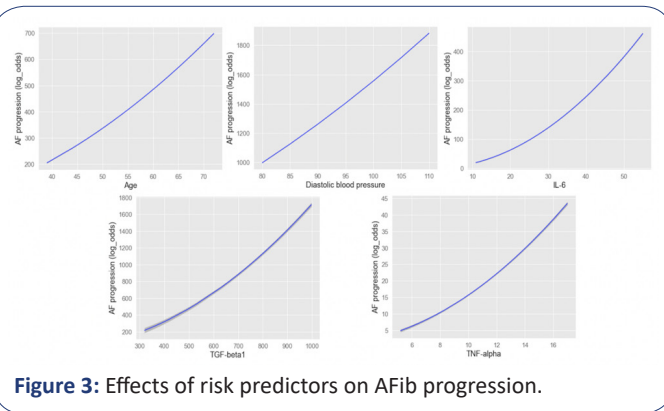
Table 5: Some results of hyperparameters selection.

λ	α	Generated risk predictors
0.5	0.0 (Ridge)	No predictor (all p > 0.05)
0.05	0.25	IL-6, TGF-beta1
0.004	0.5	Age, DBP, IL-6, TNF-alpha, TGF-beta1
10.0	0.75	No predictor (all p = 1.0)
0.02	1.0 (Lasso)	IL-6, TGF-beta1

Table 6: The results according to the best hyperparameters.

	P value
LVWT	0.565
LV PWD	0.426
BMI	0.529
MI	0.074
IHD	0.253
AGE	0.017
LV EDD	0.430
CRP	0.594
TNF-alpha	0.005
TGF-beta1	0.000
IL-6	0.000
Diastolic Blood Pressure	0.025
LAV	0.126
EF	0.183

Figure 3: Effects of risk predictors on AFib progression.

These predictors serve as independent risk factors for the onset of AFib. Figure 3 shows the relationship between risk predictors and atrial fibrillation progression. The main intent of the proposed method is to weight each risk predictor based on its effect on disease progression. The weight function applies to each risk predictor using the results of the trained model. AFib progression accelerates at different rates according to each risk predictor's increasing value.

Table 7: An illustrative of AFib progression based on risk predictors weighted approach.

Age	Weight	DBP	Weight	IL-6	Weight	TNF-alpha	Weight	TNF-beta1	Weight	Total Weight	Severity Risk	Class Type
40	216	85	1124	15	36	6.6	7	420	348	2627	0.144	Paroxysmal
50	337	90	1261	20.3	64	8.3	11	510	494	2847	0.335	Persistent
60	485	95	1405	31.6	153	9.6	15	650	770	2948	0.455	Persistent
70	660	100	1557	39.9	244	11.3	20	720	930	3073	0.609	Persistent
80	862	105	1716	45.6	318	14.5	32	850	1270	3269	0.806	Permanent
90	1091	110	1883	55.0	461	15.9	38	900	1414	3634	0.962	Permanent

New dataset construction and AFib classification

The effect of risk predictors on AFib progression is discussed above. Based on that a new dataset is constructed with weighted values of predictors which are *Age, Diastolic blood pressure, IL-6, TNF-alpha, and TGF-beta1*. The constructed novel dataset will assist clinical use and further analysis of patients. Table 7 shows an illustration of AFib progression based on the proposed methodology. First, it is presented the individual weight of each risk predictor of some patients. The total weight indicates each patient's estimated individual risk of AFib progression. The less the total weight, the less the probability of disease progression, also called the severity risk. Any severity risk lower than 0.25 presents a good outcome. The new dataset performance is evaluated using some extension of receiver operating characteristics (ROC). The results are presented in Figure 4. The ROC curve for weighted predictors confirmed its good clinical performance (micro-average AUC = 0.85, macro-average AUC = 0.88). Performance micro-average metric is preferred if the possibility of class imbalance is high [47,48].

Table 8 illustrates the multiclass AFib progression classification results based on the constructed dataset according to the proposed methodology. First, it presents the classification results on the initial RIC Dataset considering all 27 clinical characteristics; then identifies risk predictors. The last row shows the classification using weighted risk predictors (DTL - SW).
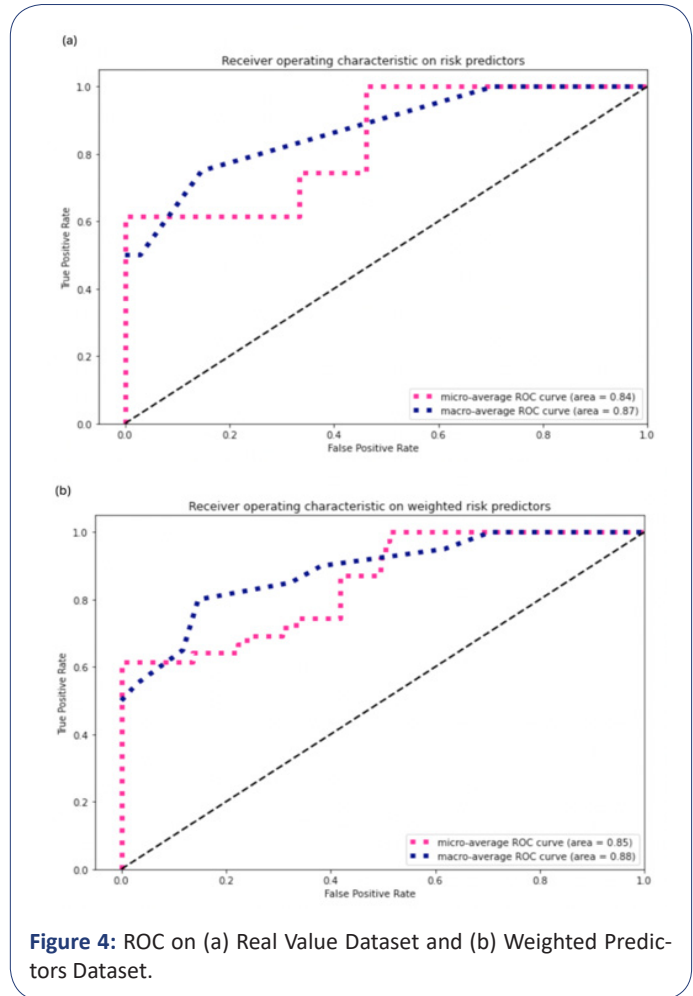

Figure 4: ROC on (a) Real Value Dataset and (b) Weighted Predictors Dataset.

Table 8: Comparison table of classification results based on selected features (accuracy).

RIC Afib Dataset	Logistic Regression	Decision Tree	Random Forest	Gaussian NB	LDA	KNN
Initial 27 features	78.04	92.14	76.92	91.22	88.46	70.73
Identified risk predictors	92.68	97.56	95.12	97.22	94.44	86.11
DTL - SW	97.93	98.36	97.77	97.56	97.22	94.44

Table 9: Comparison of suggested approach with other state-of-the-art methods (accuracy).

UCI Heart Disease Dataset	Logistic Regression	Decision Tree	Random Forest	Gaussian NB	LDA	KNN
Heart ChiSq	84.50	80.0	83.0	83.67	85.0	66.67
Heart Ref	83.0	80.0	83.5	83.67	85.0	61.67
Heart SyUn	83.33	78.33	83.0	83.0	84.67	66.67
DTL - SW	91.67	85.0	86.67	90.0	88.33	88.33

Evaluation of suggested approach on the public dataset – UCI Heart Disease

The suggested method has been experimented on a public dataset called UCI Heart Disease [50-57]. The dataset contains 303 samples and 14 clinical characteristics (features). Many of the medical dataset features are irrelevant and uninformative. This can decrease model accuracy. In real-world applications, these features need to be put in context with other patient-related data, which might be unavailable at the time of classification. We aim to identify the predictors (features) that contribute the most to the disease progression through the suggested method. Then the identified risky predictors will be used for a more accurate model building.

The riskiest predictors on UCI Heart Disease Dataset were found to be *age, cp, exang, ca, oldpeak, thal*. Then, these predictors have been weighted based on their contribution to heart disease progression. We use the weighted values for classification purposes.

In Table 9, we evaluate the results of the suggested approach with existing state-of-the-art feature selection methods [1]. The results are presented in Table 9. The simulation results show that the proposed method obtains an accuracy of 91.67% for heart disease classification and has been improved by up to 6% compared with existing approaches.

The best performing model in the [1], study was the BayesNet algorithm on Heart ChiSq features with an accuracy of 85%. We achieved an accuracy of 89% using BayesNet based on the proposed doctor-in-the-loop feature selection and weighting approach (DTL-SW).

Disussions

Being able to effectively detect Atrial fibrillation (AFib) and other related cardiovascular problems are crucial for the world. Even though the disease has existed for ages, we still miss the comprehensive framework that can deal with the solution where clinicians and machine learning can have an impact. Multiple studies show various ways of identifying predictors associated with AFib disease progression and classification. Still, very little research has been done to weight the independent risk factors of having a doctor-in-the-loop. This study a) proposed and validated a novel Doctor-In-The-Loop (DTL) machine learning system to identify and classify AFib in ischemic heart disease and arterial hypertension patients, including b) Expert-driven Pearson correlation-based feature (clinical characteristics) selection; c) Regression-based risk predictors' identification; d) risk predictors weighting based on its impact on AFib progression; and b) analysis of 257 such cases hospitalized in the department of cardiac arrhythmias of the Research Institute of Cardiology (RIC) in Armenia. Extensive computer simulations showed that the *Age, Diastolic blood pressure, IL-6, TNF-alpha, and TGF-beta1.* features have the highest effect on AFib progression. Quantitive comparisons were based on adopted accuracy, sensitivity, precision, F1-score, and ROC metrics. Patients' AFib classification showed that the newly constructed dataset based on the risk predictor weighting approach outperforms the significance of the initial RIC Dataset reaching accuracy up to 98.36%. Thus, the novel system assists the proper identification and classification of AFib patients.

The proposed method obtains an accuracy of 91.67% for heart disease classification on the well-known UCI Heart Disease Dataset. It improves by up to 6% compared to Spencer et al. method [1].

However, this study has some limitations. First, we have only used AFib data from a single center in our train, test and validation sets. Furthermore, the patients included in this study were patients who have other cardiovascular problems, i.e., ischemic heart disease and arterial hypertension. This may impact the generalizability of the proposed methodology. Therefore, the suggested methodology should be validated on other Afib datasets.

Conclusions

The identification and incorporation of weighted risk predictors based on their impact on heart disease progression through a doctor-in-the-loop approach significantly improve the identification and classification of heart disease. Thus, the presented framework has a good potential to facilitate the decision-making process in heart disease, particularly Atrial Fibrillation identification and classification. The presented method could be used as an initial screening for heart disease by helping clinicians diagnose three types of atrial fibrillation in real-time, facilitating faster decision-making and reducing costs.

Declarations

References

1. R Spencer, Thabtah F, Abdelhamid N, Thompson M. Exploring feature selection and classification methods for predicting heart disease. Digit Health. 2020; 6: 2055207620914777.

2. Müller C, Hengstmann U, Fuchs M, Kirchner M, Kleinjung F, et al. Distinguishing atrial fibrillation from sinus rhythm using commercial pulse detection systems: The non-interventional BAYathlon study. Digital Health. 2021; 7.

3. Hassan SU, Mohd Zahid MS, Abdullah TA, Husain K. Classification of cardiac arrhythmia using a convolutional neural network and bi-directional long short-term memory. Digital Health. 2022; 8.

4. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? Plos One. 2017; 12.

5. P. Ambrosino, Bachetti T, D'Anna SE, Galloway B, Bianco A, et al. Mechanisms and Clinical Implications of Endothelial Dysfunction in Arterial Hypertension. J. Cardiovasc. Dev. Dis. 2022; 9: 136.

6. D Lai, Bu Y, Su Y, Zhang X, Ma CS. "Non-Standardized Patch-Based ECG Lead Together with Deep Learning Based Algorithm for Automatic Screening of Atrial Fibrillation", in IEEE Journal of Biomedical and Health Informatics. 2020.

7. A Jalali, M Lee. "Atrial Fibrillation Prediction With Residual Network Using Sensitivity and Orthogonality Constraints", in IEEE Journal of Biomedical and Health Informatics. 2020.

8. A Rizwan, Ahmed Zoha , Mabrouk IB, Sabbour HM, Al-Sumaiti AS, et al. "A Review on the State of the Art in Atrial Fibrillation Detection Enabled by Machine Learning", in IEEE Reviews in Biomedical Engineering. 2021.

9. M Poorthuis, NR Jones, P Sherliker, R Clack, GJ de Borst, et al. "Utility of risk prediction models to detect atrial fibrillation in screened participants", in European Journal of Preventive Cardiology. 2021; 28.

10. SS Chugh, R Havmoeller, K Narayanan, D Singh, M Rienstra, et al. "Worldwide Epidemiology of Atrial Fibrillation", in A Global Burden of Disease 2010 Study. 2013.

11. G Lippi, F Sanchis-Gomar, G Cervellin. "Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge", in International Journal of Stroke. 2020.

12. F Liang, Y Wang. "Coronary heart disease and atrial fibrillation: a vicious cycle", in American Journal of Psycology of Heart and Circulatory Psychology. 2021.

13. M Poorthuis, NR Jones, P Sherliker, R Clack, GJ de Borst, et al. "Utility of risk prediction models to detect atrial fibrillation in screened participants", in European Journal of Preventive Cardiology. 2021.

14. G Boriani, M Vitolo, DA Lane, TS Potpara, GYH Lip. "Beyond the 2020 guidelines on atrial fibrillation of the European society of cardiology", in European Journal of Internal Medicine. 2021.

15. C Gutierrez, DG Blanchard. "Atrial Fibrillation: Diagnosis and Treatment", in Am Fam Physician. 2011.

16. J Heijman, D Linz, U Schotten. "Dynamics of Atrial Fibrillation Mechanisms and Comorbidities", in Annual Review of Physiology. 2020.

17. P Kieseberg, E Weippl, A Holzinger. "Trust for the Doctor-in-the-Loop", in ERCIM News of Tackling Big Data in the Life Sciences. 2016.

18. P Kieseberg, J Schantl, P Fruehwirt, E Weippl, A Holzinger. "Witnesses for the Doctor in the Loop," in International Conference on (BIH). 2015.

19. A Holzinger, I Jurisica. "Biomedical informatics: Discovering knowledge in big data", in Springer, New York. 2014.

20. A Holzinger. "Interactive machine learning for health informatics: when do we need the human-in-the-loop? ", in Brain Informatics. 2016.

21. M Nakamura, T Yamashita, A Hayakawa, T Matsumoto, ATakita, et al. "Bleeding risks associated with anticoagulant therapies after percutaneous coronary intervention in Japanese patients with ischemic heart disease complicated by atrial fibrillation: A comparative study", in Journal of Cardiology. 2021.

22. P Rasmussen, P Blanche, F Dalgaard, GH Gislason, C Torp-Pedersen, et al. "Electrical cardioversion of atrial fibrillation and the risk of brady-arrhythmic events", in American Heart Journal. 2022.

23. N Albuquerque, TL de Araujo, MV de Oliveira Lopes, TMM Moreira. "Hierarchical analysis of factors associated with hospital readmissions for coronary heart disease: A case-control study", in Journal of clinical nursing. 2020.

24. J Pearl, D Mackenzie. "The book of why. New York", in NY: Basic Books. 2018.

25. Cai, L Jiawei, W Shulin, Y Sheng. Feature selection in machine learning: A new perspective, Neurocomputing. 2018; 300: 70-79.

26. S Chowdhury, R Govindaraj, P Mayilvahanan. Optimal feature extraction and classification-oriented medical insurance prediction model: machine learning integrated with the internet of things, International Journal of Computers and Applications. 2022; 44: 278-290.

27. X Zeng, TR Martinez. Feature Weighting Using Neural Networks, Computer Science Department, Brigham Young University, Provo, Utah. 84602.

28. KJ Butt, A study of feature selection algorithms for accuracy estimation, thesis. 2012.

29. Pavithra V, Jayalakshmi V. "Comparative Study of Machine Learning Classification Techniques to Predict the Cardiovascular Diseases Using HRFLC", in 5th ICICCS. 2021.

30. A Wosiak, D Zakrzewska. "Integrating Correlation-Based Feature Selection and Clustering for Improved Cardiovascular Disease Diagnosis", in Hindawy Complexity. 2018.

31. M Saluja, D Pillai, S Sharma. "SPARC scoring model: Objective outcome prediction in critically ill COVID-19 patients", in Current Medicine Research and Practice. 2021; 11.

32. Gude F, Riveiro V, Rodríguez-Núñez N, Ricoy J, Lado-Baleato O, et al. "Development and validation of a clinical score to estimate progression to severe or critical state in COVID-19 pneumonia hospitalized patients", in Scientific Reports. 2020; 10: 19794.

33. Lang WU. "Generalized Linear Models", in Applied Multivariate Statistical Analysis and Related Topics with R. 2021.

34. H Sinkovec, G Heinze, R Blagus, A Geroldinger. "To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets", in Methodology. 2021.

35. J Ogutu, T Schulz-Streeck, HP Piepho. "Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions", in QTLMAS. 2011.

36. Comber A, Harris P. "Geographically weighted elastic net logistic regression", in J Geogr Syst. 2018; 20: 317-341.

37. Pirjatullah, Dwi Kartini, Dodon Turianto Nugrahadi, Muliadi, Andi Farmadie. "Hyperparameter Tuning using GridsearchCV on The Comparison of The Activation Function of The ELM Method to The Classification of Pneumonia in Toddlers." 2021 4th IC2IE, 2021.

    J Lorraine, D Duvenaud. "Stochastic hyperparameter optimization through hypernetworks", arXiv preprint arXiv: 1802.09419. 2018.

38. L Zhang, J Hailati, X Ma, J Liu, Z Liu, et al. "Analysis of risk factors for different subtypes of acute coronary syndrome", in Journal of International Medical Research. 2021; 49: 5.

39. K Kira, LA Rendell, A practical approach to feature selection, Proceedings of the ninth international workshop on machine learning, Morgan Kaufmann Publishers Inc. 1992; 249-256.

40. M Dialameh, MZ Jahromi. A general feature-weighting function for classification problems, Expert Systems with Applications. 2017; 72: 177-188.

41. FX Diebold, RS Mariano. "Comparing predictive accuracy", in Journal of Business & Economic Statistics, 2002.

42. P McClure. "Sensitivity and specificity", in Journal of Hand Therapy. 2001.

43. AY Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, et al. "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network", in Nature Medicine. 2019.

44. H A Haenssle, C Fink, R Schneiderbauer, F Toberer, T Buhl, et al. "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists", in Annals of Oncology. 2018.

45. Andrew J Einstein, Leslee J Shaw, Cole Hirschfeld, Michelle C Williams, Todd C Villines, et al. "International Impact of COVID-19 on the Diagnosis of Heart Disease", in Journal of the American College of Cardiology. 2021.

46. Chongwei Wu, Wei Yan, Hongtu Li, Jiaxin Li, Hongkai Wang, et al. "A classification system of day 3 human embryos using deep learning", in Biomedical Signal Processing and Control. 2021.

47. Jun Li, Qingguang Chen, Xiaojuan Hu, Pei Yuan, Longtao Cui, et al. "Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques", in International Journal of Medical Informatics. 2021.

48. D Yadav, S Pal. "Prediction of Heart Disease Using Feature Selection and Random Forest Ensemble Method", in Journal for Pharmaceutical Research Scholars. 2020.

49. A Janosi. Heart Disease Data Set.

50. F. Guerra, G Stronati. "Risk prediction models in atrial fibrillation: from theory to practice", in Journal of Preventive Cardiology. 2021.

51. Y Kim, Y Kim. "Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models", in Sustainable Cities and Society, 2022.

52. L Riyaz, MA Butt, M Zaman, O Ayob. "Heart Disease Prediction Using Machine Learning Techniques: A Quantitative Review", in Conference on Innovative Computing and Communications, 2021.

53. D Paikaray, AK Mehta. "An Extensive Approach Towards Heart Stroke Prediction Using Machine Learning with Ensemble Classifier", in International Conference on Paradigms of Communication, Computing and Data Sciences. 2022.

54. K Johnson, Johnson HE, ZhaoY, Dowe DA, Staib LH, et al. "Scoring of Coronary Artery Disease Characteristics on Coronary CT Angiograms by Using Machine Learning", in RSNA Radiology, 2019.

55. Nieuwlaat R, Prins MH, Le Heuzey JY, Vardas P, Aliot E, et al. "Prognosis, disease progression, and treatment of atrial fibrillation patients during 1 year: follow-up of the Euro Heart Survey on atrial fibrillation", in Eur Heart J. 2008.

56. S Chugh, R Havmoeller, K Narayanan, D Singh, M Rienstra, et al. "Worldwide Epidemiology of Atrial Fibrillation", in A Global Burden of Disease 2010 Study, 2010.